DOI:
10.1007/978-3-030-02628-8_7

Document Version
Peer reviewed version

Evaluation of strategies for PET motion correction - manifold learning vs. deep learning *

James R. Clough[1], Daniel R. Balfour[1], Claudia Prieto[1], Andrew J. Reader[1], Paul K. Marsden[1], Andrew P. King[1]

School of Bioengineering & Imaging Science
King's College London
London, UK
james.clough@kcl.ac.uk

Abstract. Image quality in abdominal PET is degraded by respiratory motion. In this paper we compare existing data-driven gating methods for motion correction which are based on manifold learning, with a proposed method in which a convolutional neural network learns estimated motion fields in an end-to-end manner, and then uses those estimated motion fields to motion correct the PET frames. We find that this proposed network approach is unable to outperform manifold learning methods in the literature, in terms of the image quality of the motion corrected volumes. We investigate possible explanations for this negative result and discuss the benefits of these unsupervised approaches which remain the state of the art.

Keywords: Motion estimation · Positron emission tomography · Convolutional neural network · Principal component analysis

1 Introduction

Positron emission tomography (PET) imaging is widely used for cancer management and provides information which is vital for diagnosis and monitoring of treatment. In the clinical setting PET is limited by a low signal-to-noise ratio (SNR) because high tracer doses, which increase SNR, also cause radiation exposure and cancer risk to the patient and so the dose is deliberately kept low.

Bodily motion is a further complicating factor which degrades image quality by causing blurring and image artefacts. In particular, respiratory motion is hard to avoid as it is involuntary and many minutes are required to perform a typical PET scan making breath-holding impossible for patients. One way of accounting for organ motion is to simultaneously acquire another imaging modality, such as magnetic resonance (MR) imaging, which can be used to motion correct the PET data. Simultaneous PET-MR scanners [12, 7], in which MR can be used to

* This work was supported by the Engineering and Physical Sciences Research Council under Grant EP/M009319/1 and by the Wellcome EPSRC Centre for Medical Engineering at Kings College London (WT203148/Z/16/Z).

motion correct the PET (eg. [2]), are beginning to be used clinically but make up only a small minority of existing PET scanners. Where simultaneous scans are not possible motion modelling using sequential scans can be used for motion correction (eg. [1]). However, motion modelling with sequential scans is limited in its accuracy by the assumption that the breathing patterns during the two scans do not significantly differ.

In principle an attractive solution is to estimate a respiratory signal and to use it to perform motion correction by gating acquired data based on the amplitude of that signal. This signal can be derived from the PET data or from a secondary device measuring, for example, chest position. Data driven signals are promising in that they require no secondary hardware and are directly related to the organ motion of interest, but face the challenge of extracting an accurate signal from low SNR data. A comparison of several data-driven approaches was presented in [14] which found that manifold learning methods such as PCA and Laplacian Eigenmaps performed well as methods of extracting the respiratory signal, with PCA identified as perhaps more stable in noisy conditions.

Recently, convolutional neural networks (CNN) have been shown to be capable of de-noising images taken under low-light conditions or with very short exposure times [10, 4]. Such images suffer from Poisson noise, as does PET. The use of CNNs to de-noise, or map low-dose into high-dose images in PET has also been developed recently. In [16] a residual U-net [11] architecture was used to predict full-dose images from 0.5% dose images. In [3] PET-like images were generated from CT data. It may even be possible to de-noise PET data by training a network to perform the inverse-Radon transform and output a high-quality reconstruction from raw sinogram data as is claimed in [17] although the scalability of such an approach remains a significant challenge.

CNNs have also been shown to be capable of performing non-rigid image registration [15]. Such methods can potentially be orders of magnitude faster in their run-time than traditional iterative approaches. Although training the network (i.e. learning the function required to deform each image) is slow, one forward pass (i.e. evaluating that function once to perform the registration) is fast. This approach has proved successful in cases of 2D cardiac MR [15], 3D brain MR [8] and on X-ray images [9]. Notably though, as far as we are aware, such approaches have not been applied to PET imaging, presumably because of the difficulty of dealing with low SNR and non-Gaussian noise.One might then expect that combining these two approaches could allow an appropriate CNN architecture to de-noise a PET frame and estimate the deformation required to transform it to a reference position, which would allow motion correction of a sequence of such frames.

In this paper we attempt to estimate the motion states of PET frames, by training on motion fields acquired from simultaneously acquired MR volumes. We compare a CNN-based approach with a state of the art approach based on manifold learning. We find that, despite our experimentation with various network architectures the CNN approach is unable to outperform the much simpler manifold learning approach.

2 Methods

2.1 Network Architecture

To estimate motion fields directly from time-resolved PET frames we propose a CNN which is illustrated in figure 1. The network receives two PET frames (see figure 2 for examples) as its input - the 3D volume from a reference respiratory position R, which is a fully exhaled position, and the 3D volume in question, V_t. In our experiments these volumes are both of size $48 \times 176 \times 256$ in the anterior-posterior $\times$ head-foot $\times$ left-right directions. The desired output is the set of three-dimensional motion fields, M_t which represent the deformation of the underlying anatomy from the position in V_t to the position in R, which can then be used to transform V_t into the reference motion state. The ground truth motion fields are such that $M_t(V_t) \approx R$, where $M_t(V_t)$ denotes the result of applying the transformation M_t to the volume V_t.

The output of the network is a $48 \times 176 \times 256 \times 3$ tensor, representing the required voxelwise deformation in the x, y, z directions. As a loss function we use the mean square difference between the three components of the predicted motion field vectors and the ground truth components which simply corresponds to the mean square displacement between the predicted and ground truth vectors.

2.2 Training Details

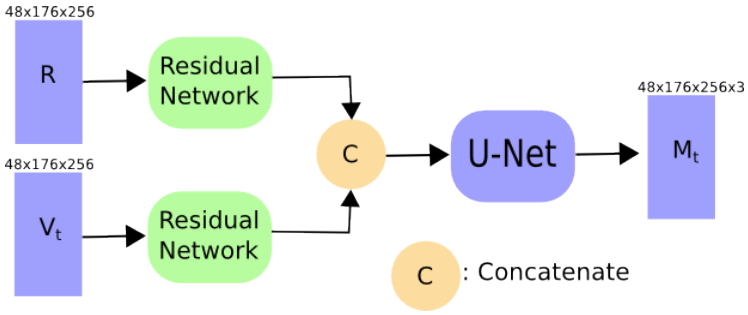
Our method was implemented in Keras[1]. The network was trained with the Adam optimiser, with a learning rate of 0.001, and with Dropout regularisation in the two convolutional layers in the lowest resolution layer of the U-net. We used a batch size of 4, the maximum allowed by our GPU memory. In all cases, the results for one subject are acquired by training the network on all PET frames from all other subjects.

3 Experiments

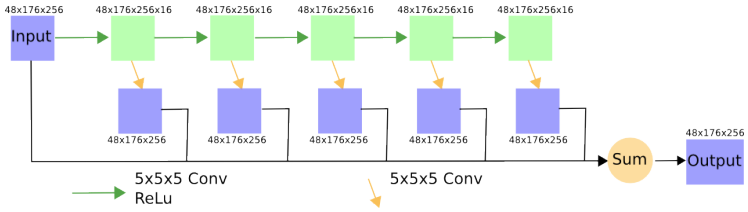
3.1 Synthetic dataset

We conducted our experiments on a highly-realistic synthetic dataset. The data consist of real MR acquisitions which are then used to create synthetic PET data, giving us a paired PET-MR dataset. The PET simulations were intended to mimic a typical ^{18}F-fluorodeoxyglucose (FDG) scan. Cardiac-gated abdominal MR scans were performed on 10 healthy volunteers, with both a high-resolution exhale breath-hold volume, and sequences of 35 low-resolution dynamic volumes acquired for each of three breathing modes, 'deep breathing', 'normal breathing' and 'fast breathing' making 105 acquired low-resolution dynamic MR volumes for each volunteer.

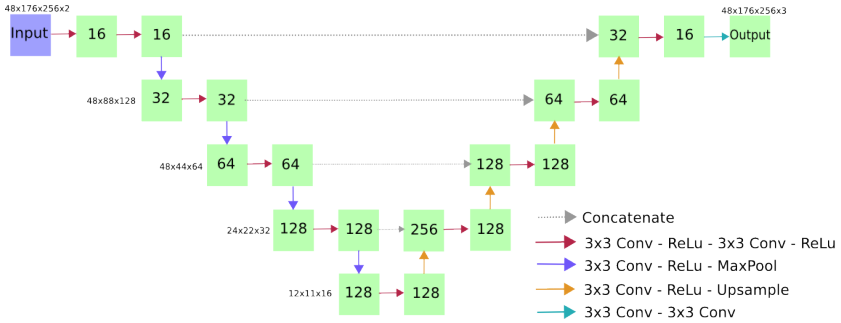
[1] https://keras.io/



(a) Overall network architecture



(b) Residual network architecture



(c) U-net network architecture

Fig. 1: Diagram of the neural network architecture used in these experiments. Both of the input volumes are first passed through a shared residual de-noising layer, before being concatenated and passed into a U-net like architecture to incorporate both local and global information into the final motion estimation.

The high-resolution volumes were segmented into anatomical regions relevant to PET emission and attenuation to create attenuation maps and FDG emission maps for each volunteer. These FDG maps were then augmented by adding artificial lesions (either one or two spherical lesions in the lungs and/or liver, of sizes between 10mm and 20mm in diameter) such that each volunteer had ten emission maps (one unmodified, four with one added lesion, and five with two added lesions).

Motion fields were extracted by performing a non-rigid registration from the high-resolution breath-hold to the low-resolution dynamic volumes. The simu-

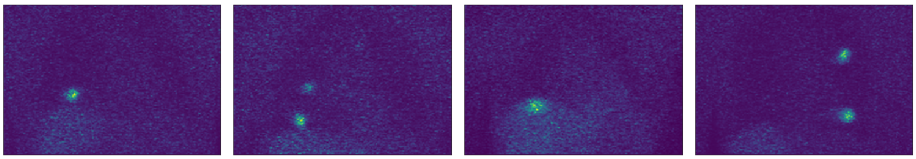


Fig. 2: Examples of typical simulated PET frames from four volunteers. The images shown are coronal sections chosen to make the lesions easily visible.

lated PET was then created by using the calculated motion fields to warp these attenuation and emission maps, from which PET sinograms were simulated and then time-resolved frames reconstructed using the ordered-subsets expectation maximisation (OSEM) reconstruction algorithm [6]. The simulations include random coincidences and scatter, with each simulated scan having a total of 50 million simulated counts and an additional 25 million random coincidences.

In total this gave us 10 volunteers each with 10 artificial lesion placements, and motion states from 3 breathing modes each with 35 acquired volumes giving a total of $10 \times 10 \times 35 \times 3 = 10500$ simulated PET frames.

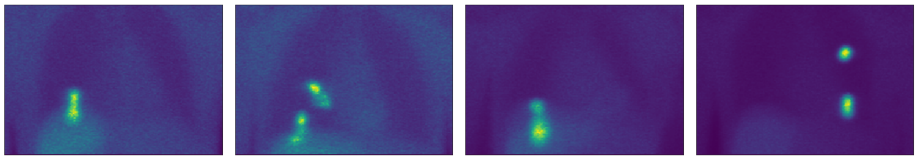
Finally, we also simulate PET acquisitions using no motion fields, producing a simulation of a theoretical acquisition in which there was no respiratory motion. This provides us with a best achievable performance for motion correction.

3.2 Comparison method: data-driven gating

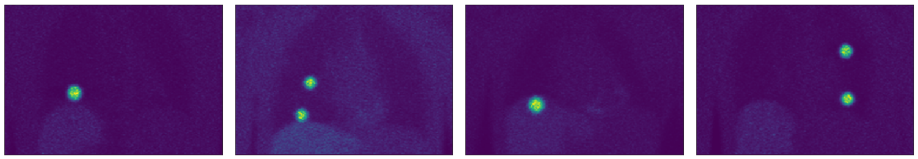
To evaluate the CNN-based motion correction approach we compare it to the unsupervised PCA-based method introduced in [13]. As implemented here, this method involves taking the Freeman-Tukey [5] transformation of the PET frames and then taking the first component of the PCA of this data (which we find always corresponds to respiratory motion) as a gating signal. The 35 PET frames for each sequence are then grouped into 5 gates using this gating signal, the data in each group aggregated, and the resulting volumes then registered to a target gate. The data from these groups are then aggregated to create the final motion corrected volume.

3.3 Assessment of corrected volume quality

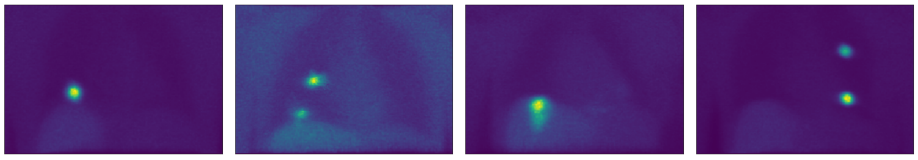
We quantitatively assessed the quality of the motion corrected volumes using the peak standardised uptake value (SUV) in the region of interest (ROI) of the lesion(s). The SUV for a voxel was found by taking a small region of interest (the voxel in question and its 6 adjacent voxels) and taking the mean intensity value across this group. The voxel within the ROI of the lesions with the highest such mean value determines the peak SUV value in that region. The lesion's ROI was defined by the lesion's position in the original segmentations used to create the simulated PET, which effectively represents a ground-truth position for the lesion. We use the peak SUV calculated in this way to compute our final evaluation metric which is a percentage SUV recovery. The peak SUV values



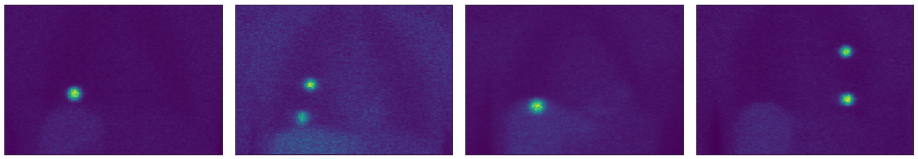
(a) Uncorrected PET volumes



(b) Motionless volumes



(c) CNN motion corrected volumes



(d) PCA motion corrected volumes

Fig. 3: Example of uncorrected volumes (top row), motionless volumes (second row), and motion corrected volumes with the CNN method (third row) and with the PCA method (fourth row), from four of the volunteers in our dataset.

for the motion correction methods assessed here are expressed as a percentage of the motionless peak SUV value. For the CNN motion correction method, we used a cross-validation scheme in which the CNN was trained on the other nine volunteers and then tested on the left-out volunteer. As is clear from table 1, while both methods of motion correction improve upon the raw, uncorrected volumes, the PCA method outperforms the CNN method on all ten volunteers. Examples of motion corrected volumes are shown in figure 3.

4 Discussion and Conclusions

Why might unsupervised methods be more powerful or more appropriate solutions for motion correction in PET than deep CNNs? Although breathing pat-

Volunteer	Uncorrected	PCA corrected	CNN corrected
1	$45.5\% \pm 11.0$	$68.6\% \pm 3.3$	$57.5\% \pm 7.4$
2	$43.4\% \pm 11.3$	$71.1\% \pm 2.5$	$57.9\% \pm 4.0$
3	$30.8\% \pm 3.6$	$60.4\% \pm 4.0$	$37.4\% \pm 6.3$
4	$32.1\% \pm 1.4$	$62.0\% \pm 7.7$	$38.1\% \pm 4.2$
5	$32.0\% \pm 5.9$	$71.9\% \pm 14.2$	$43.4\% \pm 5.8$
6	$31.3\% \pm 6.4$	$70.8\% \pm 7.4$	$39.4\% \pm 5.0$
7	$61.1\% \pm 8.9$	$77.9\% \pm 6.2$	$69.3\% \pm 5.1$
8	$58.6\% \pm 11.7$	$76.1\% \pm 5.6$	$65.0\% \pm 7.7$
9	$66.9\% \pm 4.4$	$75.3\% \pm 4.9$	$66.7\% \pm 5.6$
10	$62.5\% \pm 8.7$	$79.0\% \pm 4.7$	$70.9\% \pm 8.1$

Table 1: Percentage of peak SUV recovered using motion correction, with the gold-standard motionless volumes SUV values set to be 100% for each volunteer. Shown here are the mean and standard deviation of the peak SUV fraction over the nine artificial lesion placements, excluding the tenth case where no lesion was present.

terns between patients vary significantly, the breathing pattern for one patient over a short amount of time is often well modelled by a low-dimensional manifold [2]. This is especially true when the relevant signal in the image is highly concentrated in space, as is the case here where small lesions with high levels of FDG emission are the most important structures for motion correction. If the lesion repetitively traces out a path during respiration, and it is significantly brighter than the rest of the volume, then this signal is likely to be easily picked up by manifold learning techniques, as has been demonstrated here. More complicated organ motion which cannot be inferred from the lesion motion will not be picked up by manifold learning approaches, but importantly this kind of motion outside of the lesions will not affect the clinically relevant measurement of image quality such as the peak SUV as used here, or alternative measures such as lesion size or position.

Simple manifold learning methods may be more restrictive than CNN-based methods but in cases where the training data are very noisy, and the signal being estimated is low-dimensional, these restrictions seem to be beneficial. We note that as well as the CNN architecture described here we attempted to use several modifications which proved not to help the final image quality, including changing the sizes of the convolution kernels, numbers of layers and feature maps, and estimating joint motion fields from temporally neighbouring frames to make use of temporal correlations. We also found that, with sufficiently long training times, the CNN was able to accurately fit the training set motion fields suggesting that the problem on the test set is one of generalising to unseen subject's anatomies and breathing patterns, although further work is required to understand exactly to what extent these differences limit the final motion-corrected image quality.

While our experiments cannot demonstrate that all CNN-based methods for motion correcting PET data will struggle, they do suggest that at the very

least, when the training set is relatively small, it is very challenging to construct a CNN motion correction method for PET which approaches the performance of the state-of-the-art manifold learning methods.

Acknowledgments We would like to thank nVidia for kindly donating the Quadro P6000 GPU used in this research.

References

1. Balfour, D.R., et al.: Respiratory motion correction of PET using MR-constrained PET-PET registration. BioMedical Engineering OnLine **14**(1), 85 (2015)
2. Baumgartner, C.F., et al.: High-resolution dynamic MR imaging of the thorax for respiratory motion correction of PET using groupwise manifold alignment. Medical Image Analysis **18**(7), 939–952 (2014)
3. Ben-Cohen, A., et al.: Virtual pet images from ct data using deep convolutional networks: Initial results. In: International Workshop on Simulation and Synthesis in Medical Imaging. pp. 49–57. Springer (2017)
4. Chen, C., et al.: Learning to see in the dark. arXiv preprint arXiv:1805.01934 (2018)
5. Freeman, M.F., Tukey, J.W.: Transformations Related to the Angular and the Square Root. The Annals of Mathematical Statistics **21**(4), 607–611 (1950)
6. Hudson, H.M., Larkin, R.S.: Ordered Subsets of Projection Data. IEEE transactions on medical imaging **13**(4), 601–609 (1994)
7. Judenhofer, M.S., et al.: Simultaneous PET-MRI: A new approach for functional and morphological imaging. Nature Medicine **14**(4), 459–465 (2008)
8. Li, H., Fan, Y.: Non-Rigid Image Registration Using Self-Supervised Fully Convolutional Networks without Training Data. arXiv preprint arXiv:1801.04012 (2018)
9. Miao, S., et al.: A CNN Regression Approach for Real-Time 2D/3D Registration. IEEE Transactions on Medical Imaging **35**(5), 1352–1363 (2016)
10. Remez, T., et al.: Deep Convolutional Denoising of Low-Light Images (2017), http://arxiv.org/abs/1701.01687
11. Ronneberger, O., et al.: U-net: Convolutional networks for biomedical image segmentation. In: International Conference on Medical image computing and computer-assisted intervention. pp. 234–241. Springer (2015)
12. Shao, Y., et al.: Simultaneous pet and mr imaging. Physics in Medicine & Biology **42**(10), 1965 (1997)
13. Thielemans, K., et al.: Device-less gating for PET/CT using PCA. IEEE Nuclear Science Symposium Conference Record pp. 3904–3910 (2011)
14. Thielemans, K., et al.: Comparison of different methods for data-driven respiratory gating of PET data. IEEE Nuclear Science Symposium Conference Record pp. 3–6 (2013)
15. de Vos, B.D., et al.: End-to-end unsupervised deformable image registration with a convolutional neural network. In: Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support, pp. 204–212. Springer (2017)
16. Xu, J., et al.: 200x low-dose pet reconstruction using deep learning. arXiv preprint arXiv:1712.04119 (2017)
17. Zhu, B., et al.: Image reconstruction by domain transform manifold learning. Nature Publishing Group **555**(7697), 487–492 (2017)